10

Counts and Counted Fractions

		Page
Introduction		10-1
10A.	Relative importance: total counts	10-1
10B.	Looking effectively at counts: logarithms	10-3
10C.	Looking effectively at counts: square roots	10-7
10D.	Relative importance for counted fractions	10-8
10E.	Three matched scales for counted fractions	10-10
10F.	Rats, a family of examples	10-12
10G.	Some examples where careful expression clearly pays off	10-17
10H.	How far have we come?	10-26

*****

10K.	How to work with square roots and logarithms	10-28
10L.	Counted fractions defined by a variable cutting point	10-33
10M.	Comparison of simple and refined ways of expressing fractions	10-37

*****

10P.	IEDA processor features	10-43

*****

10Y.	Problems	10-44

plurality $f(1-f)$
f-root $\sqrt{2f} - \sqrt{2(1-f)}$

Location of tables and plots for 10

Number	Page
1	10-6
2	10-9
3	10-13
4	10-14
5	10-20
6	10-21
7	10-23
8	10-24
9	10-25
10	10-31
11	10-32
12	10-34
13	10-36
14	10-40
15	10-42
16	10-45
17	10-46
18	10-47
19	10-49

10

Counts and Counted Fractions

We often get our data by counting. Sometimes the total count is meaningful: How many planets revolve about the Sun? Or about other nearby stars? Here our attention centers on the count. In other situations what matters is a counted fraction: Out of a national sample of 1541 college students, 329 were found wearing beards. Here our attention centers on 329/1541 = .214 = 21.4%.

Either of these kinds of data -- total count or counted fraction -- tends to have special characteristics -- tends to be usefully treated in certain simple ways. Since counts occur so frequently, it will help us to be clear what these ways for dealing with them are and why they are likely to help us. To do this is our next task.

10A. Relative importance: total counts

When we have total counts, we usually want to choose a way of putting them into numbers that reflects what we really think about them.

What does such a sentence mean? It can mean many things, the simplest of which is merely that changes we regard as roughly equally "important" are expressed by roughly equal changes in how we have expressed our counts. Indeed, since our ability to compare the intuitive sizes of small changes is usually larger than our ability to make similar comparisons among large changes, we are likely to use fairly small changes in judging rough equality of change.

Consider two areas of land of roughly equal size -- we will call them counties for convenience. If one contains 10 wolves one year but 12 wolves the next, we would regard this change as a noticeable but far from large increase. If the other contains 1000 wolves one year but 1002 the next, we would regard this change as quite trivial. The fact that the difference in counts is exactly two wolves in both cases is true but quite misleading. (If you do not like to think about wolves, think about the same numbers for Boy Scout troops, Baltimore orioles, airports, Presbyterian churches, parks, or cider mills.)

What if the second county has 1000 wolves one year, but 1200 the next? If these wolves come in large, irregular packs we may be tempted to feel that the change from 1000 to 1200 is no more important than the change from 10 to 12. If we feel this way, then we could usefully speak of a "20% change" in both cases.

We are, however, more likely to feel that a change from 1000 to 1200 is considerably "larger" than a change from 10 to 12. (Particularly if, at this season of the year, wolves run in pairs.) If this happens, whatever increase over 1000 that we judge to be about as "important" as a change from 10 to 12 will have to be a change to a count that is noticeably more than 1002 and noticeably less than 1200. In many cases, a change from 1000 to 1020, which falls in this range, is wisely judged to be about as "important" as a change from 10 to 12.

In the first case our intuition matches the use of "% change". In the second case it matches nothing with which we are familiar. As we consider first a change from 10 to 12, then a "similar" change from 1000 to 1020, the change in count increases from 2 to 20 while the change in % decreases from 20% to 2%. In a sense, we want to go "halfway" between "number change" and "% change".

On these intuitive foundations, how do we erect good ways of expressing counts? What ways do we reach? (We shall leave the first of these questions to the interested reader, concentrating upon the second.)

10B. Looking effectively at counts: logarithms.

Let us come to details, beginning with the less likely case where ratios tell the story.

To say that changes in % tell us what interests us most is to say that ratios are what matter. (Whether or not counts are in question, such a clutch for ratios seems to be the typical economist's view in most situations.) To say this -- and to have it roughly true -- is a long step forward, but it is no more than a single step: It lets us express changes, but our original task was to express the counts themselves -- and to do this in such a way that differences between the numbers used will express, reasonably closely, the intuitive importance of changes in count.

Again we can turn to logs. Equal %s of change mean equal ratios -- before/after -- and hence equal logs

of these ratios. Since the log of a ratio is the difference of the logs, equal logs of ratios means equal differences of logs.

Specifically, in our first wolf example:

$$\begin{aligned} \log 12 &= 1.07918 \\ \log 10 &= \underline{1.00000} \\ \text{difference} &= .07918 \end{aligned}$$

$$\begin{aligned} \log 1200 &= 3.07918 \\ \log 1000 &= \underline{3.00000} \\ \text{difference} &= .07918 \end{aligned}$$

so that, in these circumstances, expressing counts by their logs can give results corresponding to our intuition.

Only one difficulty needs a little attention. What if the count of wolves falls to 0? The log of zero is not a useful number. We need to do something to avoid this kind of trouble. The easiest thing to do seems to be to express our count as the number given by

$$\log(\text{total count} + \text{free start})$$

where "free start" is a reasonably chosen small number, the same small number being used for all total counts that arise in a specific problem or a specific area.

Such choices as:

free start = 0.1

free start = 0.25

free start = 1

free start = 3

free start = 10

are each sometimes useful. In the absence of other guidance, using "free start = 0.25" when counts are "well behaved" but "free start = 1" when counts are "irregular" is likely to work fairly well.

Using such a free start distorts our differences a little, but not enough to make us uncomfortable. In the wolf example, for instance, the change from 10 to 12 now gives the same numerical difference as a change from 1000 to 1182. (I doubt if any of our intuitions is finely enough tuned to tell which is more closely matched to the change from 10 to 12, the change from 1000 to 1182 or the change from 1000 to 1200.)

Table **1** gives -- in one way or another -- two-decimal values for any or all of these choices of "free start" and any count up to 99. Thus log 23.25 is given directly as 1.37, while log(51+1) is given as the sum of 1.71 (the value for 51 and "free start = 0.25") and 0.01 (for the *), namely 1.72.

For counts above 99, the critical table of two-decimal logarithms (table 31 on page 2-64) should be used.

table 1 of chapter 10

Logarithmic Expressions for Counts

(For counts of 100 or more use table 23 on page 2-51)

a) TABLE, FROM 0 TO 99, OF log(count + 0.25). (*used in connection with (b) below.)

(units digit of count)

	0	1	2	3	4	5	6	7	8	9
0	-.60	0.10	0.35	0.51	0.63	0.72	0.80	0.86	0.92	0.97
1	1.01	1.05	1.09	1.12	1.15	1.18	1.21	1.24	1.26	1.28
2	1.31	1.33	1.35	1.37	1.38	1.40	1.42	1.44	1.45	1.47
3	1.48*	1.49*	1.51*	1.52*	1.53*	1.55*	1.56*	1.57*	1.58*	1.59*
4	1.60*	1.61*	1.62*	1.63*	1.64*	1.65*	1.66*	1.67*	1.68*	1.69*
5	1.70*	1.71*	1.72	1.73	1.73*	1.74*	1.75*	1.76	1.77	1.77*
6	1.78*	1.79	1.79*	1.80*	1.81	1.81*	1.82*	1.83	1.83*	1.84*
7	1.85	1.85*	1.86	1.86*	1.87	1.88*	1.88*	1.89	1.89*	1.90
8	1.90*	1.91	1.92	1.92	1.93	1.93	1.94	1.94	1.95	1.95
9	1.96	1.96	1.96*	1.97	1.97*	1.98	1.98*	1.99	1.99*	2.00

b) TABLE, FROM 0 TO 29, OF log(count + 1). (For counts of 30 to 99 use portion (a), increasing *ed values by 0.01.)

	0	1	2	3	4	5	6	7	8	9
0	0.00	0.30	0.48	0.60	0.70	0.78	0.85	0.90	0.95	1.00
1	1.04	1.08	1.11	1.15	1.18	1.20	1.23	1.25	1.28	1.30
2	1.32	1.34	1.36	1.38	1.40	1.41	1.43	1.45	1.46	1.48

c) TABLE, FROM 0 TO 29, OF log(count + 0.1). (For counts of 30 to 99 use portion (a), without change.)

	0	1	2	3	4	5	6	7	8	9
0	-1.00	0.04	0.32	0.49	0.61	0.71	0.79	0.85	0.91	0.96
1	1.00	1.05	1.08	1.12	1.15	1.18	1.21	1.23	1.26	1.28
2	1.30	1.32	1.34	1.36	1.38	1.40	1.42	1.43	1.45	1.46

d) NOTES:
(1) The accuracy of this table is usually unnecessarily great.
(2) For log(count + 3) enter portion (b) with "count + 2".
(3) For log(count + 10) enter portion (b) with "count + 9".

e) EXAMPLES:

count	log(count+0.25)	log(count+1)	log(count+0.1)	log(count+3)	log(count+10)
27	1.44	1.45	1.43	1.48	1.57
43	1.63	1.64	1.63	1.66	1.72
72	1.86	1.86	1.86	1.87	1.91

The precision provided by table 1 is, if anything, excessive. That provided by table 31 of Chapter 2 will be quite adequate in all ordinary circumstances.

10C. Looking effectively at counts: square roots.

So much, for the moment, for what we may reasonably do when equal ratios reflect equally important changes. What are we to do when, as is noticeably more common, we find a change from 1000 to 1020 a reasonable match for a change from 10 to 12?

We have already noted that what we want to do is to go "halfway" to the use of ratios, which means "halfway" to the use of logarithms. In dealing with the scale of powers, as in Graph 30 of Chapter 2, we saw that logs behave like zero powers, so that "halfway" from count to log count is likely to be the 0.5 power, corresponding to using $\sqrt{\text{count}}$. So let us try square roots.

For our wolf numbers we have:

$$\begin{aligned} \sqrt{12} &= 3.4641 \\ \sqrt{10} &= \underline{3.1623} \\ \text{difference} &= 0.3018 \end{aligned}$$

$$\begin{aligned} \sqrt{1020} &= 31.9374 \\ \sqrt{1000} &= \underline{31.6228} \\ \text{difference} &= 0.3146 \end{aligned}$$

The match is surely as close as our intuition can require.

We are thus quite free to use $\sqrt{\text{count}}$ -- or any fixed multiple of it, such as $2\sqrt{\text{count}}$ -- as a routine way of expressing counts. Or we could use something like

$$\sqrt{\text{count}} + \sqrt{1 + \text{count}}$$

if we preferred such a slightly odd-appearing expression. In fact, for reasons mainly beyond the scope of our present discussion we often do prefer just such a combination, even going further by modifying slightly a few of its values (for very small counts).

Accordingly, table 2 presents the values of such an expression for counts of 0 to 199. (For larger counts, $2\sqrt{\text{count}}$, easily obtainable from an ordinary square-root table, will work well.)

10D. Relative importance for counted fractions.

Let us now turn to counted fractions. What can we say about the relative importance of changes in them? Their usual expression is in %. We often say that 20% of all accidents have a certain type of cause. Or that 95% of a certain population can read a newspaper. Consider trying to change some simple aspect of people: can they read a newspaper, do they wear beards, do they buy "Stistix" breakfast food. If 41% can read now, a campaign to raise this figure by 1% to 42% (or in many cases to lower it by 1% to 40%) is likely to be no great task. But if only 1% -- or as many as 97% -- can read, to raise this to 2% -- or 98% -- is likely to require much effort.

table 2 of chapter 10

Values of $\sqrt{\text{count}} + \sqrt{1 + \text{count}}$

(units digit of count)

	0	1	2	3	4	5	6	7	8	9
0	1.0	2.4	3.1	3.7	4.2	4.7	5.1	5.5	5.8	6.2
1	6.5	6.8	7.1	7.3	7.6	7.9	8.1	8.4	8.6	8.8
2	9.0	9.2	9.4	9.6	9.9	10.1	10.3	10.4	10.6	10.8
3	11.0	11.2	11.4	11.6	11.7	11.9	12.1	12.2	12.4	12.6
4	12.7	12.9	13.0	13.2	13.3	13.5	13.6	13.8	13.9	14.1
5	14.2	14.4	14.5	14.6	14.8	14.9	15.0	15.2	15.3	15.4
6	15.6	15.7	15 8	15.9	16.1	16.2	16.3	16.4	16.6	16.7
7	16.8	16.9	17.0	17.1	17.3	17.4	17.5	17.6	17.7	17.8
8	17.9	18.1	18.2	18.3	18.4	18.5	18.6	18.7	18.8	18.9
9	19.0	19.1	19.2	19.3	19.4	19.5	19.6	19.7	19.8	19.9
10	20.0	20.1	20.2	20.3	20.4	20.5	20.6	20.7	20.8	20.9
11	21.0	21.1	21.2	21.3	21.4	21.5	21.6	21.7	21.8	21.9
12	22.0	22.0	22.1	22.2	22.3	22.4	22.5	22.6	22.7	22.8
13	22.8	22.9	23.0	23.1	23.2	23.3	23.4	23.5	23.5	23.6
14	23.7	23.8	23.9	24.0	24.0	24.1	24.2	24.3	24.4	24.5
15	24.5	24.6	24.7	24.8	24.9	24.9	25.0	25.1	25.2	25.3
16	25.3	25.4	25.5	25.6	25.7	25.7	25.8	25.9	26.0	26.0
17	26.1	26.2	26.2	26.3	26.4	26.5	26.6	26.6	26.7	26.8
18	26.9	26.9	27.0	27.1	27.2	27.2	27.3	27.4	27.5	27.5
19	27.6	27.7	27.7	27.8	27.9	28.0	28.0	28.1	28.2	28.2

Notes: (1) For count $\geq$ 200, the use of $2\sqrt{\text{count}}$ will be quite satisfactory.

(2) Slight modifications of the first few values may be inserted in later versions.

Examples: Count = 37, expression = 12.2.

Count = 165, expression = 25.7.

Except in determining the outcome of elections, a 1% change is a much greater thing near 0% or 100% than it is near 50%. If we are to express counted fractions in a way that reflects this, we must s t r e t c h the ends of our scale in comparison with the middle. How is it wise to do this?

First, an important remark: Treating "those who are" and "those who aren't" symmetrically is often very useful -- and sometimes important. This means two things:

- choosing scales with a value of 0 at 50%,
- arranging matters so that a swap of f with 1-f, where f is the fraction concerned, will change the sign, but not the size, of our expression of f.

While we are at it, since $f = \frac{1}{2}$ is to be scored as 0, we may as well plan to have a variety of choices that agree, not only exactly for $f = \frac{1}{2}$ but quite closely for values around $f = \frac{1}{2}$. This is a matter of convenience. It also allows ever-repeated emphasis on how similar all the expressions are "in the middle".

Matching near 50% does not cost us anything, because whether we analyze a given set of numbers or a set of numbers each exactly three times as large is almost certain to make no essential difference -- once the results are to be translated to a common, generally understandable scale.

10E. <u>Three matched scales for counted fractions</u>.

With these guideposts, what can we do that will be simple and effective? If we were not going to stretch the ends of the scale, the simplest measure that makes $f = \frac{1}{2}$ score 0 is the "plurality"

$$(f) - (1-f)$$

which is the difference between the fraction of those that are and the fraction of those that aren't.

Another reasonable name for this expression would be folded fraction. Clearly, if we fold any other expression of a fraction -- by subtracting the given function of those who aren't from the same function of those who are -- we will make $f = \frac{1}{2}$ score zero and make the score of 63% the negative of the score of 37%, whatever expression we chose to begin with. Indeed, if we are to have the sort of symmetry around 50% we seek, we can confine our attention to folded expressions. (Proof?)

While a number of more complicated expressions have been -- and may well continue to be -- popular, we can usually do quite well by folding some very simple expression. By analogy with all our other inquiries into good ways to express data, we find it natural to begin with a square root and a logarithm. We can use the square root -- or the log -- of any multiple of f. If we are to match folded fractions near 50%, it turns out that we should fold $\sqrt{2f}$ and $\frac{1}{2}\log_e f = 1.15 \log_{10} f$ (here $\log_e$ indicates the so-called natural or Napierian logarithm -- the logarithm to the base e = 2.71828...).

We need names for these expressions -- both formal names and working names. Clearly <u>folded roots</u> and <u>folded logs</u> are both formal and descriptive, while <u>f-roots</u> (perhaps pronounced "froots") and <u>f-logs</u> (perhaps pronounced "flogs") are convenient. Specifically, then, we plan to consider

$$\text{f-root} = \sqrt{2f} - \sqrt{2(1-f)}$$

$$\begin{aligned}\text{f-log} &= \tfrac{1}{2}\log_e f - \tfrac{1}{2}\log_e(1-f)\\ &= 1.15 \log_{10} f - 1.15 \log_{10}(1-f)\\ &= 1.15 \log_{10}(f/(1-f))\end{aligned}$$

as alternatives to

$$\text{plurality} = f - (1-f)$$

Table 3 sets out two-decimal values of pluralities, f-roots, and f-logs for every % from 10 to 90 and at closer intervals for smaller and larger %'s. Notice that -- to the accuracy given, which is usually sufficient -- the three expressions agree exactly for fractions between 38% and 62%. Unless we have to deal with fractions below 20% or above 80% it is likely to make only trivial differences which of these expressions we use for our counted fractions (so long as we are consistent). f-roots and f-logs were brought in to stretch the ends of the scale. If the ends of the scale are not used, it is most convenient and satisfactory that f-roots and f-logs behave very, very much like pluralities.

As always it is nice to look at changes of expression graphically as well as in numbers. In this situation, where we think and write of "stretching the tails", it is unusually desirable to compare the different scales side by side and watch -- to the extent that the feeling can be given -- the tails stretch. Plot 4 is intended to do just this.

10F. Rats, a family of examples.

Rats, wild or laboratory, are often a matter of human concern. Data about rats has to be expressed numerically for analysis. Diverse kinds of rat data deserve to be expressed in diverse ways. A review of some of these may help to put in perspective a variety of questions about how to express data -- counted as well as measured -- for analysis.

Young laboratory rats grow; by how fast they do or do not grow they can reveal the value of food supplements or the possibly deleterious effects of experimental drugs. Weight in grams (or even ounces) is often a good way to express rat size in growth studies, though it would seem that for young enough

table 3 of chapter 10

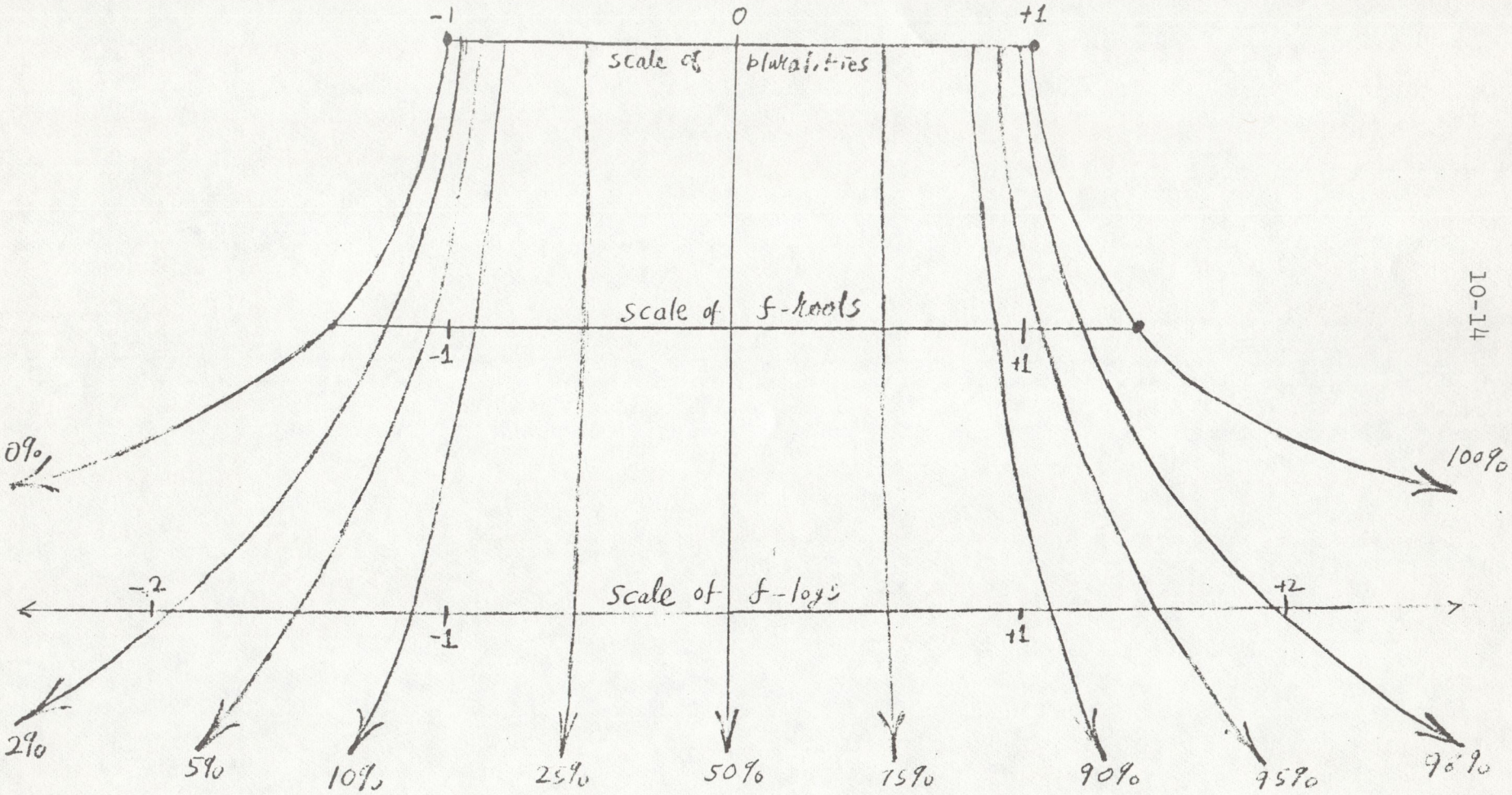
Pluralities, Folded Roots, Folded Logarithms --

Alternative Expressions for Counted Fractions

(Take sign of answer from head of column giving %)

+	Plur.	F-root	F-log	-	+	Plur.	F-root	F-log	-
50%	use	.00	use	50%	85%	.70	.77	.86	15%
51	→	.02	←	49	86	.72	.79	.91	14 *
52		.04		48	87	.74	.81	.95	13
53		.06		47	88	.76	.84	1.00	12
54		.08		46	89	.78	.87	1.05	11
55%	use	.10	use	45%	90.0%	.80	.89	1.10	10.0%
56	→	.12	←	44	90.5	.81	.91	1.13	9.5 *
57		.14		43	91	.82	.92	1.16	8
58		.14		42	91.5	.83	.94	1.19	8.5
59		.18		41	92	.84	.96	1.22	8
60%	use	.20	use	40%	92.5%	.85	.97	1.26	7.5%
61	→	.22	←	39	93	.86	.99	1.29	7 *
62		.24		38	93.5	.87	1.01	1.33	6.5
63	.26	.26	.27	37	94	.88	1.02	1.37	6
64	.28	.28	.29	36	94.5	.89	1.04	1.42	5.5
65%	.30	.30	.31	35%	95.0%	.90	1.06	1.47	5.0%
66	.32	.32	.33	34	95.5	.91	1.08	1.53	4.5
67	.34	.35	.35	33	96	.92	1.10	1.59	4
68	.36	.37	.38	32	96.5	.93	1.12	1.65	3.5
69	.38	.39	.40	31	97	.94	1.15	1.74	3
70%	.40	.41	.42	30%	97.2%	.94	1.16	1.77	2.8%
71	.42	.43	.45	29	97.4	.95	1.17	1.81	2.6
72	.44	.45	.47	28	97.6	.95	1.18	1.85	2.4
73	.46	.47	.50	27	97.8	.91	1.19	1.90	2.2
74	.48	.50	.52	26	98.0	.96	1.20	1.95	2.0
75%	.50	.52	.55	25%	98.2%	.96	1.21	2.00	1.8%
76	.52	.54	.58	24	98.4	.97	1.22	2.06	1.6
77	.54	.56	.60	23	98.6	.97	1.24	2.13	1.4
78	.56	.59	.63	22	98.8	.98	1.25	2.21	1.2
79	.58	.61	.66	21	99.0	.98	1.27	2.30	1.0
80%	.60	.63	.69	20%	99.2%	.98	1.28	2.41	0.8%
81	.62	.66	.72	19	99.4	.99	1.30	2.55	0.6
82	.64	.68	.76	18	99.6	.99	1.32	2.76	0.4
83	.66	.71	.79	17	99.8	1.00	1.35	3.11	0.2
84	.68	.73	.83	16	100.0%	1.00	1.41	*	0.0

Plot 4



Picture of the ends being stretched while the center is left undisturbed

rats -- rats that are still growing exponentially, growing by the same fraction each week -- it may be well to analyze the log of weight.

Rats run mazes, or can be trained to do so. Their performance can be measured in two ways: How rapidly do they complete their passage through the maze? How many "errors" do they make -- how often do they choose a dead-end branch? Both these measures often have to face the question of what to do with rats that never got all the way through the maze on a given trial -- perhaps because they "froze" in place at the start. An analysis in terms of seconds to complete the maze is in trouble. If a few rats take an "infinite" time -- or at least a longer time than the experimenter is willing to wait -- to complete the maze, how are their "times" to be included in the analysis? (We are now surely educated to the point where we are likely to respond "Use midmeans!" as one way out, but we may have come far enough to realize that it is even better not to h a v e to use midmeans.) The answer is simple: Look at the speed of maze-running, rather than at the time taken to run the maze. A rat that takes 12.5 seconds to run the maze has a speed of 0.08 mazes/second. A rat that refuses to run takes infinite time -- and makes zero speed. A few zero speeds do little to distort our calculations, getting over almost all difficulty. For rat maze running, then, reciprocals of the naive numbers help.

Rats also have parasites -- fleas, mites, and the like. That they do is not a trivial matter, since some of these parasites are essential to the circulation -- and transmission to humans -- of important rat-borne diseases, even bubonic plague. Thus counts of rat fleas -- or of rat mites -- can

have both scientific and very practical interest. How should they be expressed for analysis?

Fleas and mites behave quite differently. A well-gorged flea drops off the rat he -- or she -- has been feeding upon. When again hungry, this flea joins the next convenient rat. As a result, flea counts vary from one rat to another of a group, but not unusually wildly, and a square-root measure -- such as that in table 2 above, for example -- seems to work quite well as a basis for analysis.

Mites are different. They never willingly leave a rat. Rather they make a home and raise a family -- generation after generation. As a result, mite counts vary quite widely indeed, demanding a more vigorously modified expression. With mite counts it seems desirable to use a logarithmic expression, say one of those in table 1 above.

Rats can be killed, anesthetized, or visibly affected by drugs in other ways. Accordingly, such effects can be used to assess -- in biological terms -- the potency of various drugs or of various samples containing different concentrations of the same drug. (Mice are cheaper, and more often used.) To assess potency in this way, we must face the variability inherent in both rats and circumstances.

There is no such thing as a sharp, critical line between doses with no effect and doses that affect every individual rat. Experiments must be done in terms of the fraction of rats that are affected. This fraction needs to be expressed in a wisely chosen way. If we are to use one of the three ways of expression given above in table 3, we will usually do best with folded logarithms -- f-logs. (Most

practice still uses a somewhat less end-stretching scale, but its advantages are usually unproven.)

10G. Some examples where careful expression clearly pays off.

Being careful in how one expresses counts or counted fractions gives us better analyses in many, many instances. It is not always easy to be sure that this is happening to us, often the improvement is not obvious. If the data contains no two portions whose analysis we can compare, it is hard to be sure that we have indeed gained by care in choosing the expression to be analyzed. Even if one portion can be checked against another in which everything is very nearly the same, we are likely to learn little. In such a situation, the effect of a choice of expression on the two portions will be almost the same, consequently the comparison of the two for any one expression is likely to be very similar to that for any other.

To learn from internal comparisons -- from comparisons within one set of data, we need a situation where we can compare the behavior of one portion with that of another where:

- our attention is directed toward differences so that satisfactory agreement means similar differences.
- the numbers whose differences are formed -- and found to agree -- are smaller in one portion, larger in the other.

The simplest such situation, of course, is a two-way table in which both factors show substantial effects.

Thinking about the choice of form of expression -- thinking about wolves, or a substitute -- led us to consider what small changes were "similar". This is essentially a two-way table approach, although we did not say so -- and

likely did not think so. (In the wolf example, for instance, the two-way table is a 2x2 with all combinations of two years and two counties. Looking at data to see how much careful choice of expression helps -- looking at hopefully comparable differences -- also leads us to a two-way analysis approach. The similarity is far from accidental.

Let us start with a simple 2x2 example. Under the headline "Protestants shift support to Kennedy" the 2 February '62 Washington Post reported the following Gallup poll results

	Protestants		Catholics	
	Nov. '60	Jan. '62	Nov. '60	Jan. '62
Kennedy	38%	59%	78%	89%
Nixon	62%	41%	22%	11%

From an engineering (i.e. practical political) viewpoint, George Gallup was correct in saying that "one of the major -- reasons for the President's popularity has been his success in allaying anti-Catholic sentiment while not losing the support of his fellow Catholics."

But we might want to try to understand (numerically) what is really going on. We can be very sure that this situation demands scale stretching near the ends. (It is surely easier to change 38% to 42% than 89% to 93%.) So let us go to f-logs, and look at the same data again (with f = fraction for Kennedy). The f-logs are:

	Protestants	Catholics	Difference
Nov. '60	-.24	+.63	+.87
Jan. '62	+.18	+1.05	+.87
Change	up .42	up .42	

so that the Protestant and Catholic changes are very similar and mutually supporting.

Understanding of what happened to John F. Kennedy's popularity between November 1960 and January 1962 -- a scientific matter -- is much better served by the f-log analysis. We can say: "JFK's popularity improved by +0.42 on the f-log scale; you would have got this same answer had you asked only Catholics or only Protestants. It is this constancy -- present in f-logs, but not in % or pluralities -- that shows forth the advantages of our choice of f-logs.

In 1951, Neiswander reported counts on the white grub population in certain areas of turf. Eight years earlier the field in question had been divided into blocks, eight of which will be considered here. Each block was divided into plots, treated differently. After the eight-year span, the amounts of the insecticide lead arsenate remaining from each treatment differed. Table 5 gives the counts and their elementary analysis, which is rather obviously unsatisfactory. Table 6 gives the square-root expressed counts -- obtained using table 2 -- and their elementary analysis, which is clearly much more satisfactory, both as to the relative size of residuals and as to the pattern of their distribution. (We shall set the showing that each of these is true as a problem. See 10Y.)

table 5 of chapter 10

Neiswander's Counts of White Grubs in Turf Eight Years After Lead Arsenate
(Treatments: L = least remaining, ML, M, MH, H = most remaining)

	-(Data in raw counts)-					----------(Its elementary analysis)----------					
Block	L	ML	M	MH	H	L	ML	M	MH	H	(*)
1:	16	5	1	0	1	1.3	-1.7	-0.8	-0.4	+1.7	-1.2
2:	10	3	1	1	0	-3.1	-2.1	+0.8	+2.2	+2.3	-2.8
3:	10	6	0	2	1	-3.9	+0.1	-1.0	+2.4	+2.5	-2.0
4:	7	6	0	0	0	-5.7	+1.3	0.2	+1.6	+2.7	-3.2
5:	14	5	6	4	1	-2.1	-3.1	+2.8	+2.2	+0.3	+0.2
6:	26	10	6	2	0	+7.1	-0.9	0.0	-2.6	-3.5	+3.0
7:	24	14	6	1	0	+4.9	+2.9	-0.2	-3.8	-3.7	+3.2
8:	20	14	4	3	1	+1.5	+3.5	-1.6	-1.2	-2.1	+2.6
					(*)	+10.1	+2.1	-2.8	-4.2	-5.3	5.8

(*) Stripe adjustments (based on means).

Source: Neiswander, C. R. 1951. J. Econ. Entomol. 44: 221- (as quoted by Bliss and Calhoun).

table 6 of chapter 10

Data of Table 5, Expressed According to Table 2

	-(Data in square root form)-					----------(Its elementary analysis)----------					
Block	L	ML	M	MH	H	L	ML	M	MH	H	(*)
1	8.1	4.7	2.4	1.0	2.4	+0.7	-0.4	-0.4	-1.1	+1.2	-0.5
2	6.5	3.7	2.4	2.4	1.0	-0.4	-0.9	+0.1	+0.8	+0.3	-1.0
3	6.5	5.1	1.0	3.1	2.4	-0.9	0.0	-1.8	+1.0	+1.2	-0.5
4	5.5	5.1	1.0	1.0	1.0	-0.9	1.0	-0.8	-0.1	+0.8	-1.5
5	7.6	4.7	5.1	4.2	2.4	-0.9	-1.5	+1.2	+1.0	-0.1	+0.6
6	10.3	6.5	5.1	3.1	1.0	+1.4	0.1	+0.8	-0.5	-1.7	+1.0
7	9.9	7.6	5.1	2.4	1.0	+1.0	+1.0	+0.8	-1.2	-1.7	+1.0
8	9.0	7.6	4.2	3.7	2.4	-0.1	+0.8	-0.3	-0.1	-0.5	+1.2
					(*)	+3.7	+1.4	-0.9	-1.6	-2.5	4.2

(*) Stripe adjustments (based on means).

In the summer of 1950, C. B. Williams conducted trials of three light traps for catching insects. These trials were in a small piece of woodland (about 10 acres) on the famed Rothamsted Experimental Station (the first agricultural experiment station in the world). Table 7 sets out the broad outlines of the results.

(Since only one trap could be used at each location on a single night, it was necessary to have a detailed, balanced structure of traps and locations within three-day periods. The effects associated with these details were smaller than any of those in table 7, and need not concern us for our present purpose.)

As the one cell of the elementary analysis not left to the reader shows, the two-way analysis of this data in this form of expression is, to put it bluntly, pretty horrible. Table 8 shows the effect of using log(count+1) which is much more nearly satisfactory.

If we rearrange the 15 row adjustments of table 8 as a 5x3 table, we find the analysis shown in table 9. Here, not only are the separate effects of Date-Group and Position moderately distinct, but we can see an appearance (in the residuals) of Position A to go both up and down (with time) more rapidly than the others and for Position C to lag behind.

As Williams put in in the opening of his paper:

> "Our previous work on trapping insects by light, carried out for over eight years, had shown (Williams, 1937 and 1940) that the difference(s) between catches in different nights or in different traps are of a

table 8 of chapter 10

Williams's Counts (from Table 7; expressed as log(count+1) as per table 1(b))

Date -Gr.**	Pos'n	Trap 1	Trap 2	Trap 3	1	2	3	(*)
16-18	A	1.25	2.00	1.98	+.03	+.18	-.22	-.17
"	B	1.61	1.46	1.98	+.47	-.30	-.16	-.23
"	C	1.04	1.67	2.31	-.11	-.08	.18	-.24
19-21	A	1.40	2.29	2.14	+.02	+.31	-.32	-.01
"	B	1.45	1.94	2.55	.00	-.11	+.12	+.06
"	C	1.32	2.39	3.15	-.44	+.03	+.41	+.37
22-24	A	1.63	2.55	2.68	-.13	+.19	.06	+.37
"	B	1.56	2.17	2.51	.00	+.01	-.03	+.17
"	C	1.28	2.34	2.62	-.34	+.18	+.02	+.17
25-27	A	1.15	1.08	2.50	+.10	-.57	+.47	-.34
"	B	1.40	2.06	2.20	+.04	+.10	-.14	-.03
"	C	1.64	2.20	2.40	+.09	+.05	-.13	+.16
28-30	A	1.04	1.36	2.04	+.09	-.19	+.11	-.44
"	B	1.20	1.83	2.34	+.01	-.08	+.07	-.10
"	C	1.50	2.27	2.83	-.18	-.01	+.18	+.29
				(*)	-.53	.07	.45	1.92

(*) Adjustments, by means.

** All dates are August 1950.

table 9 of chapter 10
Two-way Analysis of the 15 = 5x3 Row Means from table 8

	-(Old row adjustments)-			--------(Elem. anal.)----------			
Date-Group	Pos. A	Pos. B	Pos. C	A	B	C	(*)
16-18	-.17	-.23	-.24	+.16	.01	-.18	-.21
19-21	-.01	+.06	+.37	-.03	-.05	+.08	+.14
22-24	+.37	+.17	+.17	.25	-.04	-.22	+.24
25-27	-.34	-.03	+.16	-.15	+.07	+.08	-.07
28-30	-.44	-.10	+.29	-.24	+.01	+.22	-.08
			(*)	-.12	-.03	+.15	.00

(*) Adjustments, by means.

geometric nature, and not arithmetic or additative. Thus one trap catches on an average, say, twice as many insects per night as another trap--not 193 more insects per night.

"Under these circumstances it is not correct to add catches obtained under different conditions and to use either the total or the arithmetic mean for the purpose of comparison. ...simplest... to convert the catch per night into a logarithm. ..."

When the structure of the data is such that we can compare the effectiveness of different forms of expression both clearly and sharply, the use of raw counts or raw counted fractions rarely proves better than the other choices we have illustrated.

10H. How far have we come?

This chapter has dealt with the effective expression of counts and of counted fractions.

We have been led to using either square root or logarithmic expressions for counts, usually of one of the forms

$$\sqrt{\text{count}} + \sqrt{1+\text{count}}$$

$$\log(\text{count} + \text{free start})$$

where "free start" is likely to be 0.25 or 1, but may be either larger or smaller.

We have been led to using either folded roots or folded logarithms for counted fractions, specifically

$$\text{f-root} = \sqrt{2f} - \sqrt{2-2f}$$

$$\text{f-log} = 1.15 \log_{10} f - 1.15 \log_{10}(1-f)$$

where f is a fraction ($0 \leq f \leq 1$). (These expressions are sometimes useful for fractions obtained in other ways than by counting.)

In order to see directly whether a choice of a form of expression for counts or fractions is helping us, we need data of at least a certain structure. However, sound choice of form of expression helps us in many situations where we cannot show directly that it is helping us.

All methods of assessing the usefulness of a choice of form of expression as a means of:

- facilitating looking deeper into the data, or
- expressing results in forms likely to be more widely applicable

seem to come down to making the elementary analysis of some two-way table -- actual or imagined -- more effective and reasonable.

*****

10K. How to work with square roots and logarithms.

One problem with the use of square roots is working in the desert. Given paper, pencil, and NO tables, what is the poor analyst to do? If he wants to use square roots, is his position helpless?

Some of us were taught how to "do square root" as part of arithmetic. Fewer of us remember how to do it. None of us should bother.

There is an easy way to obtain square-roots -- at least better and better approximations to square roots -- that is worth remembering. It depends on starting with a trial value and calculating

$$(\text{new trial}) = \frac{1}{2}\left\{(\text{trial}) + \frac{(\text{number})}{(\text{trial})}\right\}$$

as many times as are needed to make the answer settle down well enough for our purposes.

The largest count among the examples of 8G is 1396. The worst starting value I care to contemplate is 10. Taking this as the first "(trial)", we find, successively:

$$(\text{next trial}) = \frac{1}{2}\left(10 + \frac{1396}{10}\right) = \frac{1}{2}\,(10+139.6) = 75^-$$

$$(\text{next trial}) = \frac{1}{2}\left(75 + \frac{1396}{75}\right) = \frac{1}{2}\,(75+18.6) = 47^-$$

$$(\text{next trial}) = \frac{1}{2}\left(47 + \frac{1396}{47}\right) = 37.8(\text{error})$$

$$(\text{next trial}) = \frac{1}{2}\left(37.8 + \frac{1396}{37.8}\right) = 37.366$$

$$(\text{next trial}) = \frac{1}{2}\left(37.366 + \frac{1396}{37.366}\right) = 37.363084$$

(The error in the third step was accidental, not intentional. It was left in to show how even major errors will be taken care of in this process -- so long as the arithmetic in one step -- the last step -- is correct.)

The square root of 1396 is, to six decimals, 36.363083. We have thus come close -- far closer than we are ever likely to need to for practical data analysis.

A more reasonable start would have been 40, leading to

$$(\text{next trial}) = \frac{1}{2}\left(40 + \frac{1396}{40}\right) = \frac{1}{2}\,(40+34.9) = 37.45$$

$$(\text{next trial}) = \frac{1}{2}\left(37.45 + \frac{1396}{37.45}\right) = \frac{1}{2}\,(37.45+37.28) = 37.36$$

Here these two steps give us more accuracy than we will need.

So, some of you want to ask me, what about logs -- how can we use logs in the desert? It is easy enough to do quite well -- if you can do two times two enough times.

The first 10 powers of 2 are easy to write down:

$$2$$

$$2\text{x}2 = 4$$

$$2\text{x}4 = 8$$

$$2\text{x}8 = 16$$

$$2\text{x}16 = 32$$

$$2\text{x}32 = 64$$

$$2\text{x}64 = 128$$

$$2\text{x}128 = 256$$

$$2\text{x}256 = 512$$

$$2\text{x}512 = 1024$$

Now $2^{10} = 1024$ is close to $1000 = 10^3$, so its logarithm (to base 10) is close to 3. Hence log 2 is close to $\frac{1}{10}(3) = 0.3$. This gives us the following approximate log table:

Number:	2	4	8	16	32	64	128	256	512	1024
Log:	.3	.6	.9	1.2	1.5	1.8	2.1	2.4	2.7	3.0

If we look at 2, 4, 8, 1.6, 3.2, 6.4, 1.28, 2.56, 5.12 and 10.24, and put these in order we have the approximate log table given in table **10**. As the last column of that table shows, the error in this calculation is negligible for most purposes of data analysis.

Any one of us can reconstruct this table -- any time, anywhere -- and if one-decimal logarithms are good enough, as they are for so many purposes, we can just use this table as a critical table giving all numbers ≥ 1.28 but < 1.6 a $\log_{10}$ of 0.1, etc. (This shifts our logarithms down an average of about 0.05, but such shifts rarely disturb our analyses.)

Finally, one more point. We are likely to use 1- and 2-decimal square roots and logarithms. We can save space, time and effort -- and make fewer errors -- if we shed the decimal points while doing the arithmetic. Table 9 should have looked like table **11**. The only reason it did not was to avoid any possible confusion in the viewer -- about meaning.

Table 11 both avoids confusion while doing the arithmetic and makes it easier for a viewer to do detective work on the numbers.

table 10 of chapter 10

Approximate 1-decimal Logarithm Table

Based on the Powers of 2

Number	Approximate $\log_{10}$	(*)
1	0.0	(.000)
1.28	0.1	(.107)
1.6	0.2	(.204)
2	0.3	(.301)
2.56	0.4	(.408)
3.2	0.5	(.505)
4	0.6	(.602)
5.12	0.7	(.709)
6.4	0.8	(.806)
8	0.9	(.903)

(*) Actual 3-decimal value of $\log_{10}$.

table 11 of chapter 10

Table 9 as it Should Have Looked

(All entries are $100 \cdot \log_{10}$(something))

Date-Group	Pos. A	Pos. B	Pos. C	A	B	C	(*)
16-18	-17	-23	-24	+16	+01	-18	-21
19-21	-01	+06	+37	-03	-05	+08	+14
22-24	+37	+17	+17	+25	-04	-22	+24
25-27	-34	-03	+16	-15	+07	+08	-07
28-30	-44	-10	+29	-24	01	+22	-08
			(*)	-12	-03	+15	00

(*) Adjustments, by means.

Notes: (1) Leading zeroes NOT suppressed in the interests of (a) detecting failures to write down one or the other digit and (b) having a first digit to look at.

10L. Counted fractions defined by a variable cutting point.

Not infrequently we want to deal with counted fractions defined by a cutting point that we can choose to set in more than one location. Thus, when soldiers about to be released from the army after World War II were asked about their plans to return to their previous employer (The American Soldier 1950, vol. 4, page 629), their answers were classified as: definite plans, tentative plans, considering returning, not considering returning. It would be reasonable to divide those more definitely returning from those less definitely returning at any one of three cutting points. Table 12 gives the percentages corresponding to each of these three cuts under various circumstances. Each of these is a counted fraction, so that we naturally begin by converting them to f-roots. Once this is done in (b) of table 12, we are ready to see whether the three cutting points lead us to similar conclusions.

(c) of table 12 shows the changes in f-root associated with each step of increase in duration of previous employment. In 5 out of 6 cases there is good agreement among the three cutting points. (The group of September separatees with two-five years previous employment seems to show irregular behavior.)

To the extent that all three cutting points show the same behavior, we are free:

- to use the results for any one of them. (In this example, if one cut were to be used, that between "plans" and "considering" would perhaps be most natural.)

table **12** of chapter 10

% and f-roots for Each of Three Cutting Points as to Plans to Return to Previous Employer

Cutting points →		**Definite plans**	**Tentative plans**	**Consider return**	Not consider
a) % STRONGER PLAN TO RETURN					
Separ.	Yrs. empl.				
July	**<1**	12	21	41	
	1-2	22	33	51	
	2-5	31	43	60	
	>5	49	65	79	
Sept	**<1**	18	40	57	
	1-2	35	59	78	
	2-5	47	64	77	
	>5	64	80	93	
b) F-ROOTS FOR STRONGER PLAN TO RETURN (THAN CUTTING POINT)					
July	**<1**	-.84	-.61	-.18	
	1-2	-.59	-.35	.02	
	2-5	-.39	-.14	.20	
	>5	-.02	.30	.61	
Sept	**<1**	-.68	-.20	.14	
	1-2	-.30	.18	.59	
	2-5	-.06	.28	.56	
	>5	.28	.63	.99	
c) CHANGE IN F-ROOT					
July	**<1 to 1-2**	.25	.26	.20	
Sept	"	.38	.38	.35	
July	**1-2 to 2-5**	.20	.21	.18	
Sept	"	.24	.10	-.03	
July	**2-5 to >5**	.37	.44	.41	
Sept	"	.34	.35	.45	

d) EXAMPLE. Of July separatee with < 1 year employment, 12% had definite plans, 21% had definite plans or tentative plans, and 41% had definite plans or tentative plans or were considering returning.

- to use the mean, median, midmean, etc. of the values for the various cuts.

Table 13 offers one summarization. The results for the more extreme cuts have been approximately adjusted to match those for the middle cut and the mean of those adjusted values not apparently discrepant formed. (Clearly we have been somewhat subjective in judging apparent discrepancy.) As (b) of table 13 shows, the results for the two months of separation do have a consistent description -- one in which the increase in intention to return to previous employment increases by about 0.4 (in f-roots) between July and September.

Analysis of the same 8x3 table by midmeans (see problem 8) gives increases in f-roots of .33,.53,.28,.29, showing good agreement except for the group with 1-2 years employment where the increase is noticeably larger. Why? We can always use such an elementary analysis to combine the results for different cuts.

table 13 of chapter 10

One Form of Summarization of the f-roots of Table 12

(a) ADJUSTMENT AND COMBINATION

Month	Duration	(1)	(2)	(3)	Mean*
July	**<1**	-.59	-.61	-.53	-.58
	1-2	-.34	-.35	-.33	-.34
	2-5	-.14	-.14	-.15	-.14
	>5	.23	.30	.26	+.26
Sept	**<1**	(-.33)	-.20	-.21	-.20
	1-2	.05	(.18)	(.24)	.05
	2-5	.29	.28	(.21)	+.28
	>5	.63	.63	.64	+.63

(1) f-root for left-hand cut +.25 (July) or +.35 (September).

(2) f-root for middle cut.

(3) f-root for right-hand cut -.35 (both months).

(*) Mean of values not in parentheses (which were applied somewhat subjectively for September).

(b) RESULTS IN 4x2 FORM

Duration	July	September	(Change)
<1	-.58	-.20	(+.38)
1-2	-.34	+.05	(+.39)
2-5	-.14	+.28	(+.42)
>5	+.26	+.63	(+.37)

10M. Comparison of simple and refined ways of expressing fractions.

We have confined ourselves to three very simple ways of expressing fractions:

- pluralities = folded fractions
- f-roots = folded square roots
- f-logs = folded logarithms.

We can form and fold any fractional power we choose. If we want to match behavior near 50% we should use

$$\frac{2^\lambda}{2\lambda}\left(f^\lambda - (1-f)^\lambda\right)$$

which is often most conveniently computed as

$$\frac{(2f)^\lambda - (2-2f)^\lambda}{2\lambda}$$

for whatever λ we may choose. We will call what we get this way "folded λ-powers". Thus f-roots are folded 0.5-powers and f-logs play the role of folded zero-powers. (Recall from Chapter 2 that -- as a function, properly scaled -- the log fills in neatly between small positive powers and small negative powers.) Folded powers give us a large and convenient family of ways of expressing fractions.

Three more complicated -- or perhaps refined -- ways of expressing fractions are frequently used -- or have been recently suggested. Some of the reasons for their choice lie outside our scope, but we can surely look at their values, comparing them with one another and with appropriate folded fractional powers.

One is known under such names as "the angular transformation", "anglits", "sinits", or just "angles". Its mathematical expression, when matched near 50% is

$$\theta = (\text{the angle whose sine is } \sqrt{f}) - (\text{a right angle})$$

with all angles measured in radians (1 radian = 57.3$^-$ degrees).

A second way of expression is known under such names as "probits" and "normits", and is associated with the "inverse of the cumulative Gaussian distribution". (If you know what this is well and good -- if you do not, don't worry.) This choice is very widely used, but has no really simple mathematical formula. The simplest we can do, if we want to match near 50% is to use:

$$\text{matched normit} = \sqrt{\frac{2}{\pi}} \cdot x$$

where

$$\text{fraction} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2} du$$

(Again if you recognize these expressions, fine -- if you do not, relax and go on.)

A third refined way of expression has recently been suggested by Cox and Snell. It can be expressed in terms of an incomplete beta function of fractional order. (What can we helpfully say further?) Again we are free to look at its numerical values and see how it behaves.

Finally we notice that a fair amount of use is given to "logits", which can be defined by

$$\text{fraction} = \frac{1}{1 + e^{-(\text{logit})}}$$

This is, however, nothing new to us, since

$$(\text{logit}) \equiv 2\cdot(\text{f-log})$$

is an identity.

Table 14 compares pluralities, f-roots, f-logs, three carefully selected fractional powers, the matched forms of anglits, normits and logits, and the Cox-Snell suggestion. We see that each of the more refined ways of expressing a fraction can be quite closely matched by an appropriate folded power.

Looking in more detail, we see that: The Cox and Snell suggestion is matched by the folded power with $\lambda = .57$ to within ±0.01 over the entire range. Anglits are matched by the folded power with $\lambda = .41$ to within ±0.01 from 0.1% to 99.9%, and to within ±0.05 overall. Normits are matched by the folded power with $x = .14$ to within ±0.01 from 0.5% to 99.5%, and within ±0.03 from 0.1% to 99.9%. (This degree of matching is all the more striking when we note that these two have to differ by an infinitely large amount for the most extreme fractions -- 0% or 100%.)

table 14 of chapter 10

Summary Comparison of Ways of Expressing Fractions

+	Plur.	λ = .57	(1)	f-root	λ = .41	(2)	λ = .14	(3)	f-log	-
50%	.00	.00	.00	.00	.00	.00	.00	.00	.00	50%
60%	.20	.20	.20	.20	.20	.20	.20	.20	.20	40
70%	.40	.41	.41	.41	.41	.41	.42	.42	.42	30
80%	.60	.63	.63	.63	.64	.64	.67	.67	.69	20
85%	.70	.75	.74	.77	.77	.78	.83	.83	.86	15%
90%	.80	.88	.88	.89	.92	.93	1.03	1.02	1.10	10
92.5%	.85	.96	.95	.97	1.01	1.02	1.15	1.15	1.26	7.5
95%	.90	1.03	1.03	1.06	1.11	1.12	1.32	1.31	1.47	5
96%	.92	1.07	1.07	1.10	1.16	1.17	1.41	1.40	1.59	4%
97%	.94	1.10	1.11	1.15	1.21	1.22	1.51	1.50	1.74	3
98%	.96	1.14	1.15	1.20	1.28	1.29	1.65	1.64	1.95	2
99%	.98	1.20	1.20	1.27	1.37	1.37	1.87	1.86	2.30	1
99.5%	.99	1.23	1.24	1.31	1.43	1.43	2.06	2.06	2.65	0.5%
99.9%	.98	1.26	1.27	1.37	1.52	1.51	2.44	2.47	3.45	0.1
100.0%	1.00	1.30	1.29	1.41	1.62	1.57	3.57	∞	∞	0.0

(1) This is Cox and Snell's (1968?) suggestion, matched near 50%.

(2) These are anglits (which are already matched).

(3) These are matched probits or normits.

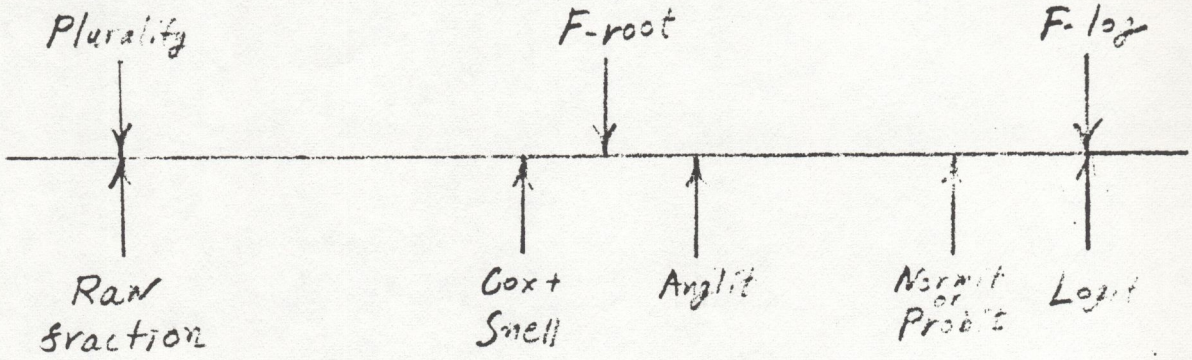
As a result, we may take the matching values of λ as good indicators of how the more refined expressions behave. Plot 15 pictures the position, and suggests why f-root and f-log seem likely to meet almost all our needs.

The Cox and Snell suggestion is new -- as yet untested by extensive experience. Its main intended uses are those for which anglits have been long and successfully used. Clearly the use of f-roots is a reasonable compromise, whichever of these two refined expressions proves the better under the test of long-continued trial.

Of course, f-logs match logits exactly. They are reasonably close to normits-probits. There may well be situations where the latter will be appreciably better than either f-logs or f-roots, but we have every reason to think that these situations will be uncommon. In most situations we expect to do quite well with one of the three simple expressions -- pluralities, f-roots or f-logs. (If we need to use some other expression, normits are a reasonable candidate, but so are folded powers for a wide variety of values of λ.)

plot 15 of chapter 10

Picture of the relationship among the simple and refined expressions for counts.



*****

10P. IEDA processor features.

In its initial form, the IEDA processor has no special capabilities for calculating log(count + free start), $\sqrt{\text{count}} + \sqrt{\text{count+1}}$, f-roots or f-logs. Those wishing to use such expressions will thus have to use an appropriate number of instructions to accomplish these operations. The following patterns should work:

log(count+1) in X(4) from data in X(3):

TRAN VAR(10) = 1 , SUM(3,10) = 4, NOW = T

TRAN LOG(4) = 4, NOW = T,

$\sqrt{\text{count}} + \sqrt{\text{1+count}}$ in X(4) from data in X(3):

TRAN VAR(10) = 1 , SUM(3,10) = 10, NOW = T

TRAN SQT(3) = 4, SQT(10) = 10, SUM(4,10) = 4, NOW = T

f-roots in X(4) from FRACTIONS in X(3):

TRAN VAR(10) = 1 , DIFF(10,3) = 10, NOW = T

TRAN SQT(10) = 10, SQT(3) = 4, DIFF(4,10) = 4, NOW = T

f-logs in X(4) from FRACTIONS in X(3):

TRAN VAR(10) = 1 , DIFF(10,3) = 10, NOW = T

TRAN LOG(10) = 10, LOG(3) = 4, DIFF(4,10) = 4, NOW = T

TRAN VAR(9) = 0.5 , PROD(4,9) = 4, NOW = T

Note: IEDA's logs are to base e so that the values of log(count + 1) thus obtained will be 2.303 times those found using logs to base 10.

*****

10Y. Problems.

1. Contrast the patterns of residuals in tables 5 and 6 to show why the elementary analysis of table 5 is less satisfactory than that of table 6.

2. Contrast the relative sizes of the residuals and the main effects in tables 5 and 6 to the same end.

3. Use IEDA to carry out the elementary analysis of table 7. Plot the observed values against the partial description.

4. Watson gives the number of lesions per half leaf found on the third and fourth leaves of tobacco plant inoculated with plant virus Hy. III. The data are given in table **16**. What form of expressing counts leads to a reasonably satisfactory analysis?

5. Watson also reports the results of an experiment in which aphids were fed for one period on an infected plant and then for a second period on a healthy plant, the response being the number of healthy plants, out of 30, later found to be infected. The data are in table **17**. What form of expression seems indicated? To what appearances does your analysis lead you?

6. Winsor and Clarke reported rather extensive results on the catch of plankton with different nets. Table **18** presents estimated numbers of plankton of various kinds caught in five horizontal hauls in which one large net (75 cm. diameter) was towed at 30 meters depth while two small nets (12.7 cm.

table 16 of chapter 10

Watson's Counts of Virus Lesions on Half-leaves

	Third leaf		Fourth leaf	
Plant #	Right	Left	Right	Left
1	56	73	32	50
2	33	32	16	35
3	47	58	22	48
4	58	48	47	50
5	72	68	16	21
6	75	56	66	83
7	40	60	29	49
8	76	101	17	39
9	179	106	23	18
10	77	63	18	25

Source: M. A. Watson 1936. Factors affecting the amount of infection obtained by Aphis transmission of the virus Hy. III. Phil. Trans. Roy. Soc. London 226: 457-489, especially table IV on page 464.

table 17 of chapter 10

Watson's Data on Number of Plants Out of 30 Infected With Plant Virus Hy. III After Aphis Feeding

Time on infected plant	Time on healthy plant					
	2 min	5 min	15 min	1 hr	6 hr	12 hr
2 min	17	20	13	18	16	23
5 min	11	13	10	14	18	15
15 min	10	6	10	9	9	18
1 hr	3	3	2	2	7	4
6 hr	3	1	7	6	6	12
12 hr	6	9	8	2	4	9

Source: M. A. Watson 1936. Factors affecting the amount of infection obtained by Aphis transmission of the virus Hy. III. Phil. Trans. Roy. Soc. London 226: 457-489, especially table XI at page 477.

table 18 of chapter 10

Winsor and Clarke's Data on the Catch of Plankton in Nets of Different Sizes

(Large = 1 net; small = 2 nets combined)

Haul	Net	Calanus finmarchus cop. V	C. finmarchus adult	Metridia Halceus adult	Centropages typicus adult	Euthemisto juv.
1	L	50500	21000	17400	55500	2490
	S	1263	617	751	2250	63
2	L	54500	15800	33000	25200	2800
	S	1560	401	1235	1260	111
3	L	71000	8950	79000	25500	2700
	S	2450	374	3520	1550	251
4	L	27800	4950	32000	19600	2420
	S	710	111	932	763	151
5	L	19800	3030	13200	33500	965
	S	642	120	324	1265	62

Source: C. P. Winsor and G. L. Clarke 1940. A statistical study of variation in the catch of plankton nets. J. Marine Research 3: 1-34, especially top of table VII at page 32.

diameter) were towed one meter deeper and shallower respectively. What form of expression leads to a reasonable analysis? What are the results of that analysis?

7. In 1942 Beale reported on the results of certain insecticidal trials. Each of six blocks were divided into 12 plots, and each of six treatments were applied to a pair of plots, selected at random, in each block. (Since Beall's report was methodological, we do not know what the six treatments were.) The response was the number of insects, of a specified kind, remaining on the plot. Table **19** presents the data. What forms of expression are plausible? What are the results of applying them, using IEDA?

8. Take the 8x3 panel (b) of table 12 and analyze by midmean differences. Find the 4x2 table corresponding to panel (b) of table 13, and the change column. Comment.

9. Find three (3) separate reasons why the analysis of table 5 is unsatisfactory.

table 19 of chapter 10

Beall's Counts of Numbers of Phlegemontius quinquemaculata

Treatment	Block 1	2	3	4	5	6
1	10, 7	20, 14	14, 12	10, 23	17, 20	14, 13
2	11, 17	21, 11	16, 14	17, 17	19, 21	7,13
3	0, 1	7, 2	3, 1	2, 1	3, 0	1, 4
4	3, 5	12, 6	4, 3	5, 5	5, 5	2, 4
5	3, 5	3, 5	3, 6	1, 1	3, 2	6, 4
6	11, 9	15, 22	15, 16	13, 10	26, 26	24, 13

Source: G. Beall 1942. The transformation of data from entomological field experiments so that the analysis of variance becomes applicable. Biometrika 32: 243-262, especially table 7 at page 246.